基于改进Mask R-CNN模型的工厂化养蚕蚕体识别与计数

何锐敏[1]，郑可锋[2]，尉钦洋[1]，张小斌[2]，张　俊[1]，朱怡航[2]，赵懿滢[2]，顾　清[2*]

（1. 嵊州陌桑高科股份有限公司，浙江绍兴 312400；2. 浙江省农业科学院数字农业研究所，浙江杭州 310021）

摘　要：精准饲喂是全龄人工饲料工厂化养蚕节本增效的核心技术之一，家蚕自动化识别与计数是实现精准饲喂的关键环节。本研究基于机器视觉系统获取工厂化养蚕过程中蚕在4龄和5龄期的数字图像，利用改进深度学习模型Mask R-CNN检测蚕体和残余饲料。通过在Mask R-CNN模型框架中加入像素重加权策略和边界框细调策略，从噪声数据中训练一个鲁棒性更好的目标检测模型，实现模型性能的优化，提高对蚕体和饲料边界的检测和分割能力。改进Mask R-CNN模型对蚕的检测和分割交并比阈值为0.5时的平均精度（Average Precision at IoU=0.5，AP50）分别为0.790和0.795，识别准确率为96.83%；对残余饲料的检测和分割*AP*50分别为0.641和0.653，识别准确率为87.71%。模型部署在NVIDIA Jetson AGX Xavier开发板上，单张图像平均检测时间为1.32 s，最长检测时间为2.05 s，运算速度可以满足养蚕盒单元在生产线上移动实时检测的要求。该研究为工厂化养蚕精准饲喂信息系统和投喂装置的研发提供了核心算法，可提高人工饲料的利用率，提升工厂化养蚕生产管理水平。

关键词：家蚕；人工饲料；精准饲喂；机器视觉；深度学习；mask　R-CNN；噪声数据

中图分类号：S126;TP18　　**文献标志码：**A　　**文章编号：**SA202201012

引用格式：何锐敏, 郑可锋, 尉钦洋, 张小斌, 张俊, 朱怡航, 赵懿滢, 顾清. 基于改进mask R-CNN模型的工厂化养蚕蚕体识别与计数[J]. 智慧农业(中英文), 2022, 4(2): 163-173.

HE Ruimin, ZHENG Kefeng, WEI Qinyang, ZHANG Xiaobin, ZHANG Jun, ZHU Yihang, ZHAO Yiying, GU Qing. Identification and counting of silkworms in factory farm using improved mask R-CNN model[J]. Smart Agriculture, 2022, 4(2): 163-173. (in Chinese with English abstract)

1　引　言

全龄人工饲料工厂化养蚕是一种全新的家蚕饲养模式，它以人工饲料代替桑叶，通过控制环境，以工业化流水线方式实现家蚕自动化大规模周年循环饲养，颠覆了传统的家蚕饲养模式，是蚕业转型升级的重要方向[1-3]。全龄人工饲料工厂化养蚕技术体系的研究取得了诸多成果，已建成投产了全龄期使用人工饲料的规模化养蚕工厂[4,5]。人工饲料的生产是工厂化养蚕的主要投入，饲料的加工和饲喂是最重要的技术环节之一。因此，严格控制饲料用量、提高饲料利用率对控制工厂化养蚕成本具有十分重要的意义。目前，工

收稿日期：2021-11-02

基金项目：浙江省重点研发计划项目（2019C02001）

作者简介：何锐敏（1964—），女，学士，研究方向为工厂化养蚕。E-mail：agri_intel@163.com

*通信作者：顾　清（1986—），男，博士，副研究员，研究方向为农业信息技术。E-mail：guq@zaas.ac.cn

厂化养蚕采用常量投喂模式,即在同一龄期中,每个养蚕盒单元中的人工饲料投喂量是相同的。然而,随着饲养的进行,由于蚕盒中的初始蚁蚕数量不同、蚕死亡、出现疑似病症人工移除、蚕爬离蚕盒等多种因素会导致蚕盒中的蚕数出现较大的差异。因此,常量投喂会导致喂食不均、饲料过多或不足等情况,最终导致饲料浪费或蚕茧质量下降。因此,精准投喂人工饲料对提高饲料利用率、降低饲养成本和提高蚕茧的质量具有重要意义。根据蚕的数量投喂人工饲料是一种有效实现蚕精准饲喂的方法,需要首先快速准确地检测每个养蚕盒单元中的蚕数,并将其转换成该蚕盒所需投喂的饲料量并传输给投喂装置,装置收到信号后即时调整出料量,达到精准饲喂目的。

图像识别中的目标检测技术可用于对蚕进行识别并计数。传统的目标检测方法多是基于目标物体边缘相关特征[6,7],虽然在特定场景下能够得到较好的检测精度和速度,但自适应性和泛化性较弱。近年来,深度学习技术被广泛应用于目标检测。基于深度学习的目标检测方法能够自适应提取不同层次的图像特征,训练得到的模型可以应用于不同场景,大幅度提升模型的精度和泛化能力[8-12]。Mask R-CNN模型是近年来较为常用的深度学习算法之一,在众多的应用场景中均取得了优秀的表现。Mask R-CNN通过使用RoIAlign改进了Faster R-CNN的RoIPooling,并且采用双线性插值算法来减小边界框回归的位置误差[13]。Mask R-CNN模型不仅具有较好的检测性能,还能对检测目标进行像素级的分割,符合本研究应用场景的需求。但Msak R-CNN在进行蚕体和残余饲料检测时,如果数据标注不清、标注区域存在其它背景、目标物体轮廓重叠和粘连等,训练数据中存在一定的噪声,会降低模型检测和掩码分割的精确性和稳定性。针对此问题,本研究利用噪声数据对Mask R-CNN模型进行改进和调整,在模型框架中加入像素重加权策略和边界框细调策略,以提升模型对蚕体和饲料边界的分割能力。

本研究利用机器视觉系统获取人工饲料工厂化养蚕过程中4龄和5龄蚕的数字图像,并利用噪声数据对深度学习模型Mask R-CNN进行改进,用于对图像中的蚕体和残余饲料进行目标分割,之后利用分割掩码输出对蚕生长发育和饲料剩余等情况进行评估,为工厂化养蚕中人工饲料精准饲喂设备和管理系统的研发提供算法支持,实现工厂化养蚕中人工饲料投喂量的精确控制,提高饲料利用效率。

2 材料和方法

2.1 数据采集和预处理

2.1.1 数据采集

数据采集地点为浙江省嵊州陌桑高科股份有限公司的工厂化养蚕人工饲料喂食车间(29°35' N,120°51' E)。家蚕品种为"中2016×日2016",是为工厂化养蚕培育的特有品种。图像获取环境温度为(25±1)℃,湿度为60%～70%。

从蚁蚕到结茧,蚕需要经历1～5个龄期,每个龄期之间会进行一次休眠(蜕皮)。每次蜕皮后,蚕体都会长大,并需要更多的食物。前3龄的蚕只需要很少的食物,约占整个生长周期所需饲料的5%,4龄期和5龄期消耗饲料占整个蚕生长周期中消耗饲料总量的95%以上。因此,本研究仅针对4龄和5龄的蚕进行识别。

图像采集设备为工业相机FLIR Blackfly S USB3,镜头型号为长步道FA3516A。相机主要规格参数为:分辨率2000万像素,定焦镜头焦距35 mm,光圈F2.8,C-Mount镜头类型、最大分辨率5472×3648、帧率18 f/s、像素大小2.4 μm。蚕盒在流水线系统上以1.5 m/s的速度水平移动,并在投喂饲料时有5 s左右的停止。蚕盒的尺寸为60cm×100cm。相机安装在养蚕盒单元停止处正上方2 m处,并在蚕盒停留间隙进行图像拍摄。相机通过USB 3.0接口连接到笔记本电脑,使用相机附带的应用程序进行控制操作。

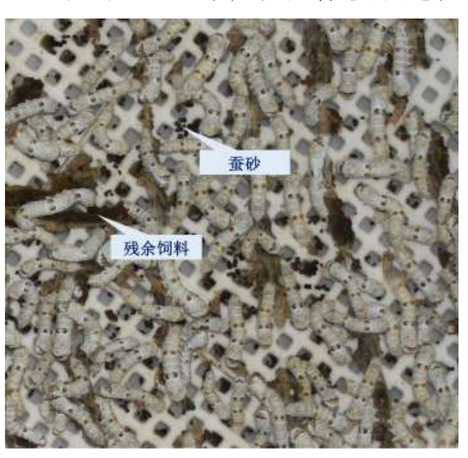
2.1.2 图像预处理

原始图像的尺寸较大,蚕体数量较多,给目

标物体标注和建模带来一定困难。为提高标注效率，统一图像尺寸，便于数据处理和后续分析，对原始图像数据进行了同尺寸裁剪，裁剪像素为2000×2000，从中挑选较为清晰的数据进行后续处理。

数据增广（Image Augmentation）可以提高图像数据的质量、扩大训练数据集的规模[14]。本研究使用旋转与翻转、亮度增强、添加噪声3种方法对原始图像进行增广。旋转与翻转是广泛使用的图像增广方法之一[14,15]，对所有图像进行90°、180°以及270°旋转和水平镜像翻转处理。由于工厂化养蚕对车间照度有严格要求，蚕室光线较暗，无法采用常规方法进行补光，原始图像的亮度偏低，所以将图像的亮度提升20%，用于弥补环境光强的不足。此外，在图像获取过程中，图像采集设备的不稳定性可能会导致随机噪声的产生。针对此，参考Zhou等[14]的方法，在原始图像中加入方差为0.01的高斯噪声，用于提升模型的鲁棒性。处理后的数据增加到训练集中，进行模型训练。

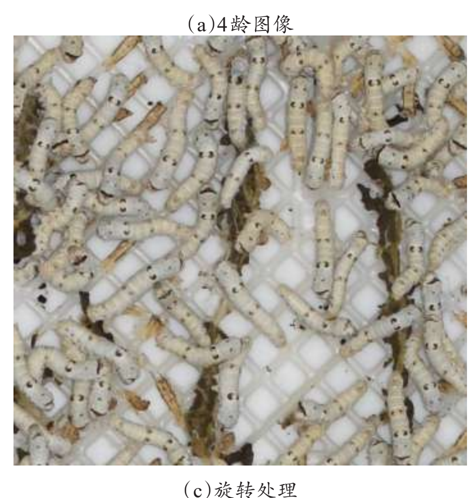
在对图像进行增广和编号后，使用开源图形界面标注工具Labelme进行图像标注。通过绘制多边形来标注蚕体和残留的饲料，标注完成后将图片保存为*. Json文件。总共选择了180张裁剪后的高质量图片进行标注，其中90张为4龄，90张为5龄。每张图像包含约150~200个蚕体标注和30~50个饲料残留标注，对图像边缘处不完整的蚕也进行标注。图1为裁剪后的4龄和5龄蚕的原始图像及不同预处理结果示例。



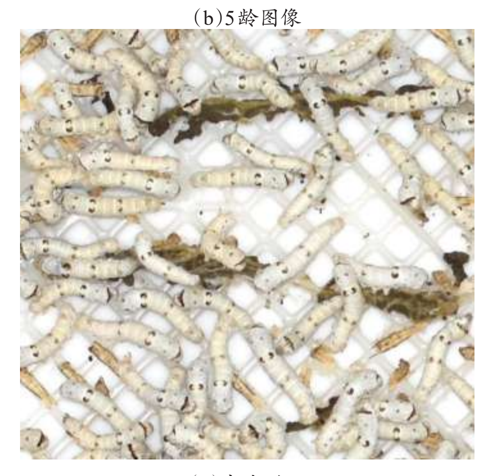
(a)4龄图像



(b)5龄图像



(c)旋转处理



(d)亮度增强


(e)噪声添加


(f)标注结果

图1 裁剪后的蚕和残余饲料原始图像及不同预处理结果

Fig. 1 Original images after cropping and different pretreatment results

2.2 Mask R-CNN

Mask R-CNN由He等[13]提出，通过在Faster R-CNN网络中添加掩码分支，在目标检测的同时实现了实例分割。该模型沿用了Faster R-CNN的思想，特征提取采用ResNet-FPN架构，另外增加了一个Mask预测分支。Mask R-CNN主要分为三个模块：Faster R-CNN、RoIAlign和全卷积网络（Fully Convolutional Networks，FCN）。其采用了与Faster R-CNN相同的两阶段检测方法。在第一阶段建立区域候选网络（Region Proposal Network，RPN），然后进行特征提取[16]。在第二阶段，Mask R-CNN引入了RoIAlign方法代替Faster R-CNN的重采样方法RoIPooling[17]。除了类别检测，Mask R-CNN会为每个候选对象输出一个二进制的分割掩码[18]。Mask R-CNN的损失函数L表示为：

$$L = L_{cls} + L_{loc} + L_{mask} \tag{1}$$

其中，L_{cls}是分类损失；L_{loc}是框回归损失；L_{mask}为掩码损失。

为了针对不同大小的家蚕进行多尺度蚕体预测，采用特征金字塔网络（Feature Pyramid Network，FPN）进行多尺度特征学习。FPN采用自上而下和自下而上的双向多尺度边框预测方法，可以将各个层级的特征进行融合，使其同时具有强语义信息和强空间信息[19]。图2所示为FPN的结构示意图。本研究中的Mask R-CNN采用ResNet50-FPN的骨干网络进行特征提取。在RPN生成过程中，anchor的大小为32、64、128、256和512，尺度为0.5、1.0和2.0。

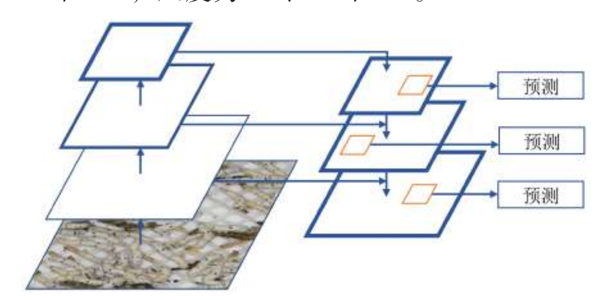


图2 特征金字塔网络结构示意图

Fig. 2 The structure of feature pyramid network (FPN)

2.3 利用噪声数据改进Mask R-CNN模型

2.3.1 原始Mask R-CNN模型存在问题及改进思路

原始Mask R-CNN模型虽然在自然图像分割上效果优越，但在分割小物体以及遮挡严重等情况下分割性能仍需改进。由于相互重叠、粘连以及其它背景物体的干扰，蚕体和残余饲料的边界可能存在模糊、难以界定的情况，导致标注数据中存在一定的错误，给模型的训练带入噪声。有噪声情况下的分割模型可能出现以下问题：①错误类别标签损坏检测器；②错误分割掩码误导模型产生不精确的掩码预测。一个过大的标注区域

将导致覆盖更多的背景区域，而过小的标注区域无法覆盖完整蚕或饲料。这些因素使得模型无法生成精准掩码；③有噪声的标注导致不稳定训练过程。当训练数据没有正确的类别标签或精确的掩码标注时，模型在训练过程中将出现不稳定的现象，损失函数不稳定将导致学习参数无法收敛至更优解。针对上述标注噪声对模型造成的影响，本研究在原始Mask R-CNN框架基础上，增加一个像素重加权策略和边界框细调策略，从噪声数据中训练一个更鲁棒的目标检测模型，实现模型性能的优化和提升。

2.3.2 主要处理流程

基于噪声数据改进的Mask R-CNN模型训练过程如下。

（1）将数据集分为训练集与元测试集，其中元测试集中的图片为人工检验后完全正确的标注数据，而训练集的标注存在噪声。

（2）使用迁移学习策略对预训练CNN模型进行微调和重训练。预训练模型使用COCO（Common Objects in Context）数据集训练。使用原始Mask R-CNN框架流程进行训练集数据训练，获得初始分类、边框回归和掩码分割结果。

（3）像素重加权策略将分类损失值作为输入，θ_1表示像素重加权策略中的参数。通过两层感知机后输出权重β作为像素的噪声水平。当像素标签不正确时，像素重加权模块输出较小的权重，而对于正确的像素标签，像素重加权模块输出的权重较大。

（4）边界框细调策略将边界框回归损失作为输入，θ_2表示边界框细调策略中的参数。通过两层感知机后输出边框位移变换参数，从而自适应地修改不精确的标注边界，并产生更准确的估计。经过边界框细调后，检测器将通过更准确的回归损失进行模型优化。

像素重加权策略和边界框细调策略的具体算法详见Xu等[20]的介绍。图3所示为基于噪声数据改进的Mask R-CNN模型用于蚕体和残余饲料检测的框架流程。将所有180张图像分成训练集、验证集、元测试集和测试集4个数据集。随机选取100张图像作为训练集，用于模型的训练。选取20张图像作为验证集，进行超参数的调试和选取。元测试集中的图片为人工检验后完全正确的标注数据，用于像素重加权和边界框细调，共有20张图像。40张图像作为测试集用于模型测试和性能评价。每个数据集中的4龄和5龄蚕图像数量均相同。

2.4 模型性能评价

预测结果可分为四类：真正例（True Positives，TP）、真负例（True Negatives，TN）、假正例（False Positives，FP）和假负例（False Negatives，FN）。TP表示被模型预测为正的正样本，TN表示被模型预测为正的负样本，FP表示被模型预测为正的负样本，FN表示被模型预测为负的正样本。在判定上述四种预测类型的归属前，需要预先定义交并比（Intersection over Union，IoU）阈值。IoU衡量检测边界与真实边界（标注边界）之间的重叠率，表示为检测结果和标注区域的重叠部分占两个区域的并集面积的比例：

$$IoU = \frac{CandidateBox \cap GroundTruth}{CandidateBox \cup GroundTruth} \tag{2}$$

其中，$CandidateBox$为检测结果边界；$GroundTruth$为标注范围边界。如果IoU值等于1，则表示预测结果与标注完全重合；IoU越接近1，预测结果越好。在本研究中，定义$IoU = 0.5$为判定预测结果的阈值，即如果IoU大于0.5，将该预测结果视为一个成功的预测，归入上述四种类别中的一种；如果小于0.5，则将其视为一个失败的预测。

在本研究中，利用准确率（Accuracy）、平均精度（Average Precision，AP）和检测速度3项指标对Mask R-CNN模型的性能进行评价。

2.4.1 准确率

准确率是正确分类的样本数占样本总数的比例[8]。

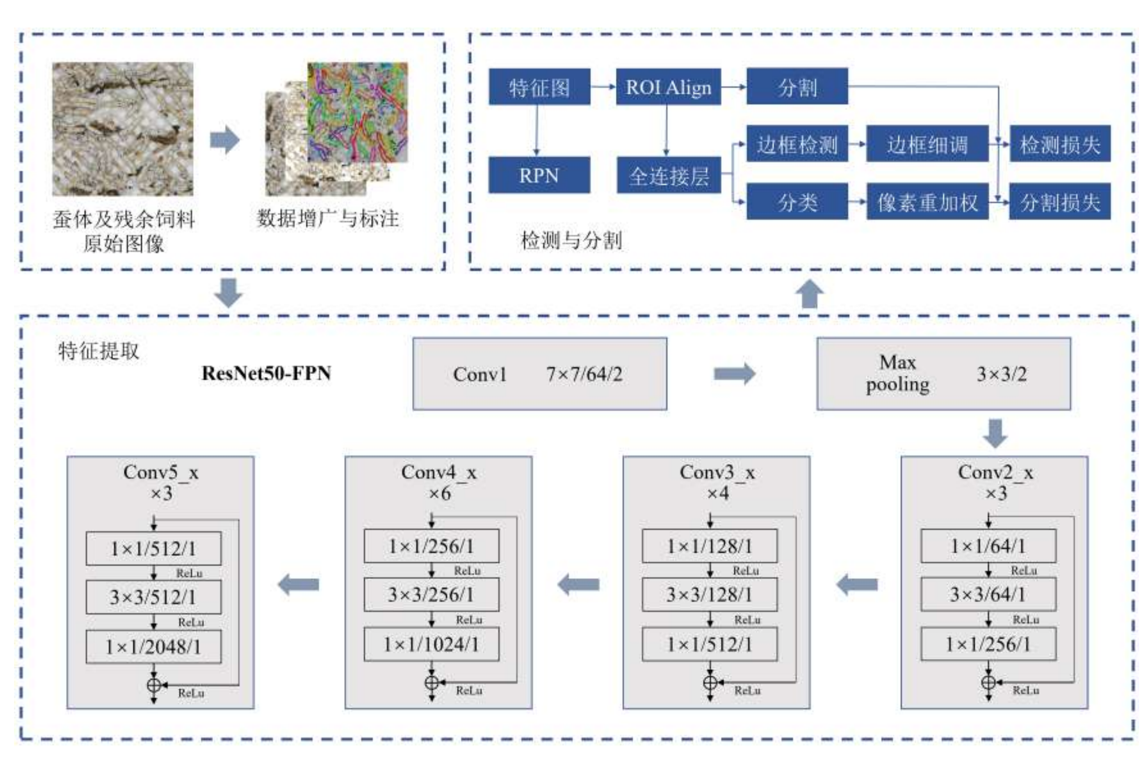


图3 基于改进Mask R-CNN模型的工厂化养蚕蚕体和残余饲料识别检测流程

Fig. 3 The workflow of the silkworm and residual feed recognition and detection based on the improved Mask R-CNN model

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

其中，$TP+TN$为正确分类的数量；$TP+TN+FP+FN$为样本总数。整个数据集的准确率是所有图像预测结果的平均值。

2.4.2 平均精度

平均精度（Precision）和召回率（Recall）是评价深度学习模型性能的常用指标，计算方式如下：

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (5)$$

其中，$TP+FP$表示目标物的预测数量；$TP+FN$为目标物的真实数量。查准率表示预测结果中正确的比例，查全率表示正确预测的目标样本占目标样本数量的比例。当Precision和Recall都较高时，模型性能较好。将Precision作为纵轴，Recall作为横轴绘制曲线（P-R曲线）。AP的定义为不同Recall值下的平均Precision值，计算方法为P-R曲线的积分[21]：

$$AP = \int_0^1 p(r)dr \quad (6)$$

其中，p表示Precision；r为Recall。AP值即为P-R曲线下方围成的面积。AP是最常用的目标检测模型性能评价指标之一。本研究中，AP在IoU阈值为0.5下计算，表示为$AP50$。

2.3.3 检测速度

运算速度是检测目标算法的重要评价指标。在本研究中，饲料投喂过程在流水线上进行，养蚕盒单元持续移动并只在饲料投放时作短暂的停止。因此，为实现不间断检测，模型运算速度必须满足一定要求。采用最大运行时间T_{max}和平均运行时间（Average Running Time，ART）两个指标对深度学习模型的运算速度进行评价[14,16]。在特定硬件配置下，T_{max}表示模型在测试集图片上检测所需的单张最长时间，ART表示模型在测试图像上执行检测过程所需的单张平均时间，s/张，描述为：

$$ART = \frac{N_t}{N_I} \quad (7)$$

其中，N_t代表模型检测所有测试图像的总运行时间，s；N_I是图像的数量，张。

3 试验与结果分析

3.1 试验设置和模型训练参数

使用迁移学习策略对预训练的模型进行微调和重训练。预训练模型使用COCO数据集训练。初始学习率为0.02，在50,000步时降低为0.0001，在70,000步时降低到0.00001。FPN在2、3、4和5残差块单元的输出上执行。研究使用开源深度学习框架PyTorch用于模型训练，编程语言为Python。试验在Ubuntu操作系统上进行，计算机配置32.0 GB内存和Intel®Core TM i7-9700K CPU @ 3.60 GHz×8处理器。在四个NVIDIA Tesla V100图形处理单元（Graphics Processing Unit，GPU）上并行训练。使用验证集数据进行模型评估和超参数调整，选取最优的参数组构建模型，再用测试集数据进行模型性能评估。模型其他初始参数见表1。

表1 蚕体和残余饲料识别模型的初始化参数

Table 1 The initialization parameters of the model for silkworm and residual feed

图像尺寸/px	批尺寸	动量	初始学习率	总步长/步	权重衰减
512×512	16	0.8	0.02	85,000	0.0005

3.2 检测与分割结果

将预测结果与标注数据进行比较，评估模型的性能。改进Mask R-CNN模型的检测结果如图4所示。其中边框表示蚕和残余饲料的识别检测结果，掩码表示分割结果。从图中可以看出，改进Mask R-CNN模型在定位目标对象方面表现良好，可以准确识别出蚕。尽管蚕蜕与蚕体较为相似，模型仍可以准确地区分两者。对于相互重叠的蚕也有较好的识别效果。



图4 基于改进Mask R-CNN模型的蚕体和残余饲料检测结果

Fig. 4 The prediction results of silkworm and residual feed based on the improved Mask R-CNN model

表2显示了Mask R-CNN模型和改进Mask R-CNN的检测和分割准确率以及*AP*值结果。Mask R-CNN模型的蚕检测和分割*AP*50分别为0.764和0.768，识别准确率为95.23%；残余饲料识别表现与蚕体相比较差，检测和分割*AP*50分别为0.602和0.611，识别准确率为85.35%。改进

Mask R-CNN 模型的蚕检测和分割 $AP50$ 分别为0.790和0.795，识别准确率为96.83%；残余饲料检测和分割 $AP50$ 分别为0.641和0.653，识别准确率为87.71%。上述结果表明，改进 Mask R-CNN 模型相比 Mask R-CNN 模型在性能上有一定的提升，并且在蚕体和残余饲料的检测和分割方面表现出良好的性能，该模型可作为工厂化养蚕人工饲料精准投喂控制系统和投喂硬件装置开发的核心算法。

表2 两种蚕体与残余饲料检测模型的性能

Table 2 The performance of the silkworm and residual feed detection models

模型	目标	准确率/%	检测 $AP50$	分割 $AP50$
Mask R-CNN	蚕	95.23	0.764	0.768
	饲料	85.35	0.602	0.611
改进 Mask R-CNN	蚕	96.83	0.790	0.795
	饲料	87.71	0.641	0.653

在检测速度方面，训练完成的改进 Mask R-CNN 模型在配置 NVIDIA Tesla V100 显卡和 i7-9700K CPU 的计算机上检测测试集图像的 ART 为 0.075 s，T_{max} 为 0.142 s。在 NVIDIA Jetson AGX Xavier 开发板上部署模型进行测试，检测 ART 为 1.32 s，T_{max} 为 2.05 s。该运算速度可以达到饲喂流水线上基于机器视觉系统的蚕体和残余饲料的实时检测要求。

3.3 不同龄期蚕分割模型表现

由于不同龄期蚕的形态特征不同，使用蚕不同生长阶段的图像数据训练的模型可能出现不同的检测表现。为比较蚕不同龄期数据集训练模型的性能，将整个数据集分为4龄和5龄两类，分别进行蚕体检测模型的训练和测试。从表3所示的模型测试结果可以看出，5龄模型比表2中完整数据模型的表现更好，而4龄模型的表现低于完整数据模型。两者相比，5龄数据集训练的模型在测试数据上比4龄模型展现出更好的性能。这表明训练图像的龄期将影响模型的检测性能。蚕龄较大的模型要好于蚕龄较小的模型，这是因为4龄的蚕体型相对较小，颜色暗黄且分布密集，而5龄的蚕具有更清晰的轮廓特征、较大的个体和更少的重叠情况。

表3 不同龄期蚕数据训练改进 Mask R-CNN 模型的蚕体检测性能

Table 3 The silkworm detection performances of the improved Mask R-CNN models using the training data of different instars

龄期	准确率/%	检测 $AP50$	分割 $AP50$
4龄	95.21	0.768	0.774
5龄	97.82	0.801	0.815

3.4 蚕体重叠对检测性能影响

生产中，蚕体之间相互重叠的情况较多，可能会影响检测结果。本节分析了改进 Mask R-CNN 模型对重叠蚕的检测性能。图5为重叠情况下的蚕的检测和分割输出实例。在测试图像中，出现重叠情况的蚕体被单独统计。测试集中共存在823条被覆盖的蚕，占总蚕数的13.4%。被重叠后显露的两段或三段蚕体被准确识别为同一条蚕的结果为正确识别，这些蚕体的总体检测准确率为95.06%，略低于整个测试数据集的检测准确率（96.83%），这说明重叠情况在一定程度上影响了检测的性能。尽管如此，改进 Mask R-CNN 对重叠蚕的检测性能依然较为理想，准确率接近于完整数据集，说明该模型对蚕体的检测性能较强，可以应对较为复杂的相互重叠情况。



(a)4龄



(b)5龄

注：红色三角形指示的蚕体被正确检测和分割，黄色三角形指示的蚕体未被模型检测

图5 蚕体重叠情况下的检测结果

Fig. 5 Detection results for overlapped silkworms

3.5 数据增广对检测性能影响

为评估数据增广对模型性能的影响，对使用完整数据建立的模型和使用剔除不同处理图像的数据集构建的模型进行了性能比较。根据表4所示，3种不同的图像增强方法对模型精度产生了不同程度的影响。其中亮度增强方法对模型性能的贡献最大。删除亮度增强处理的图像，模型的准确率降低了3.49%，检测*AP*50和分割*AP*50也有明显的下降。旋转与翻转处理对模型性能的帮助较弱，移除旋转与翻转图像后的模型准确率降低了2.04%。高斯噪声添加对模型的性能没有明显的影响。

表4 利用不同数据集训练的改进Mask R-CNN模型的蚕体检测性能

Table 4 The silkworm detection performances of the improved Mask R-CNN model trained by different datasets

数据集	准确率/%	检测 *AP*50	分割 *AP*50
完整数据集	96.83	0.790	0.795
剔除旋转与翻转处理数据	94.79	0.773	0.771
剔除亮度增强数据	93.34	0.747	0.751
剔除噪声添加数据	96.02	0.785	0.791

4 讨论与结论

4.1 讨论

人工饲料工厂化养蚕是养蚕业的技术创新，为中国养蚕业的发展提供了新的方向。根据蚕的数量饲喂人工饲料，可有效提高饲料利用率，降低成本，提高蚕茧的整体品质。蚕的准确计数还可为预测蚕茧产量、估算饲料用量、计算家蚕损失量提供参考数据，为生产决策提供支持，提高工厂化养蚕的生产管理水平。近年来，深度学习技术在目标检测中得到了越来越多的应用。基于数字图像的目标检测结果取决于目标物体的大小、格局、重叠度、图片质量和训练样本量等多种因素。在本研究中，数据类别（不同龄期、重叠蚕体）会影响检测结果，这与Tian等[22]的研究结果一致。数据增广处理能够提高深度学习模型的检测能力，这与其他研究结果相一致[14,15]。亮度增强方法对模型的性能提高具有最大的贡献。

随着整个技术体系的不断更新和完善，大规模全龄人工饲料工厂化养蚕产业化日趋成熟。本研究证实了深度学习技术在工厂化养蚕中蚕和饲料残渣检测的可行性。然而，目前还存在一些问题需要在未来的工作中解决。例如，蚕体相互重叠较多会降低模型的检测性能，后续需要进一步分析其影响程度并增加此类训练样本，以提高模型的检测和分割能力。在残余人工饲料检测方面，由于其形状不规则，表面质地多样且不均匀，以及部分被蚕体覆盖，给识别带来了较大的困难。另外，家蚕粪便的颜色和表面纹理与残留饲料相似，使得数据标注较困难，容易出错，而错误的标注又会导致识别的错误。因此，模型在残余饲料检测性能方面还存在较大的提升空间。

利用改进Mask R-CNN模型的分割掩码输出可以进一步分析蚕的大小和匀整度，以及残余人工饲料的重量估算。这些信息可用于饲料用量管理、蚕茧大小和匀整性预测等环节。此外，该技术在家蚕育种的表型分析，如蚕大小、匀整性、死亡率和生长速率计算评价等方面具有很大的应用潜力。

4.2 结论

为实现自动识别养蚕盒蚕头数，本研究提出了一种利用噪声数据改进的Mask R-CNN模型，用于家蚕和残留人工饲料的检测，为工厂化养蚕中精准饲喂管理系统和投喂装置的研发提供核心算法。

利用噪声数据改进的Mask R-CNN模型对家蚕和人工饲料残渣具有较好的检测能力，总体检测准确率分别达到96.83%和87.71%，对蚕的检测和分割*AP*50分别为0.790和0.795，对残余饲料的检测和分割*AP*50分别为0.641和0.653。

模型在NVIDIA Jetson AGX Xavier开发板上测试的*ART*为1.32 s，T_{max}为2.05 s，可以实现工

业流水线上蚕体和残余饲料的实时检测。

本研究建立的模型在精度和运算速度上可满足工业流水线上养蚕盒单元中蚕体的快速准确检测，因此可作为工厂化养蚕中人工饲料精准投喂控制信息系统和投喂装置开发的核心算法。

未来的工作重点是进一步提高模型的精度、鲁棒性和稳定性。为提高模型的性能，将引入更大规模的训练图像数据集，特别是蚕体相互重叠和粘连的样本，以及更多样的残余饲料样本。另外，下一步将选取不同结构的模型进行对比分析，对蚕体和残余饲料的图像特征提取、检测和轮廓分割作更深入的研究。

参考文献：

[1] TANAKA Y, SUDO M. Studies on the technology of artifical diet rearing for parental strains of the silkworm, 4: The relationship between the water content of artificial diets for the fifth larval instar and egg laying results[J]. Journal of Dainippon Silk Foundation, 2006, 53: 1-5.

[2] 吴亚群, 张升祥, 王洪江, 等. 家蚕不同品种对人工饲料摄食性的遗传模式[J]. 蚕业科学, 2017, 43(4): 603-609.
WU Y, ZHANG S, WANG H, et al. Inheritance pattern of feeding habit on artificial diet in different bombyx mori varieties[J]. Science of Sericulture, 2017, 43(4): 603-609.

[3] 钱秋杰, 陈伟国. 家蚕人工饲料研究与应用进展[J]. 蚕桑通报, 2016, 47(2): 11-14.
QIAN Q, CHEN W. Research and application progress of artificial diet for silkworm[J]. Bulletin of Sericulture, 2016, 47(2): 11-14.

[4] 董久鸣, 潘美良, 吴海平. 加快推进蚕桑产业转型发展的思考——巴贝工厂化养蚕的探索与启示[J]. 蚕桑通报, 2018, 49(2): 14-16.
DONG J, PAN M, WU H. Thinking on speeding up the transformation and development of sericulture Industry —The exploration and enlightenment based on the BABEI's silkworm rearing in the factory[J]. Bulletin of Sericulture, 2018, 49(2): 14-16.

[5] 王亮, 胡帅栋. 推进全龄人工饲料工厂化养蚕的巴贝模式[J]. 蚕桑通报, 2020, 51(1): 37-45.
WANG L, HU S. Babe model of promoting industrial sericulture by feeding artificial diet of full larval stage[J]. Bulletin of Sericulture, 2020, 51(1): 37-45.

[6] DOU J, LI J. Robust object detection based on deformable part model and improved scale invariant feature transform[J]. Optik-International Journal for Light and Electron Optics, 2013, 124(24): 6485-6492.

[7] HONG G S, KIM B G, HWANG Y S, et al. Fast multi-feature pedestrian detection algorithm based on histogram of oriented gradient using discrete wavelet transform[J]. Multimedia Tools and Applications, 2015, 75 (23): 1-17.

[8] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. Advances in Neural Information Processing Systems, 2012, 25: 1097-1105.

[9] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, New York, USA: IEEE, 2014.

[10] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 37(9): 1904-1916.

[11] GIRSHICK R. Fast R-CNN[C]// 2015 IEEE International Conference on Computer Vision. Piscataway, New York, USA: IEEE, 2015.

[12] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(6): 1137-1149.

[13] HE K, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[C]// IEEE International Conference on Computer Vision. Piscataway, New York, USA: IEEE, 2017.

[14] ZHOU C, HU J, XU Z, et al. A Novel greenhouse-based system for the detection and plumpness assessment of strawberry using an improved deep learning technique[J]. Frontiers in Plant Science, 2020, 11: ID 559.

[15] TIAN Y, YANG G, WANG Z, et al. Apple detection during different growth stages in orchards using the improved YOLO-V3 model[J]. Computers and Electronics in Agriculture, 2019, 157: 417-426.

[16] 张远琴, 肖德琴, 陈焕坤, 等. 基于改进Faster R-CNN的水稻稻穗检测方法[J]. 农业机械学报, 52(8): 231-240.
ZHANG Y, XIAO D, CHEN H, et al. Rice panicle detection method based on improved Faster R-CNN[J]. Transactions of the CSAM, 52(8): 231-240.

[17] WEN Q, LUO Z, CHEN R, et al. Deep learning approaches on defect detection in high resolution aerial images of insulators[J]. Sensors, 2021, 21(4): ID 1033.

[18] WJPD A, YT A, RONG L.B, et al. Detection and segmentation of overlapped fruits based on optimized mask R-CNN application in apple harvesting robot[J]. Computers and Electronics in Agriculture, 2020, 172 (6): ID 105380.

[19] M.AKHAN, ZHANG Y D, SHARIF M, et al. Pixels to classes: Intelligent learning framework for multiclass skin lesion localization and classification[J]. Computers and Electrical Engineering, 2021, 90: 1-20.
[20] XU Y, ZHU L, YANG Y, et al. Training robust object detectors from noisy category labels and imprecise bounding boxes[J]. IEEE Transactions on Image Processing, 2021, 30: 5782-5792.
[21] ZHANG Y, CHU J, LENG L, et al. Mask-refined R-CNN: A network for refining object details in instance segmentation[J]. Sensors, 2020, 20(4): ID 1010.
[22] TIAN Y, YANG G, WANG Z, et al. Instance segmentation of apple flowers using the improved mask R-CNN model[J]. Biosystems Engineering, 2020, 193: 264-278.

Identification and Counting of Silkworms in Factory Farm Using Improved Mask R-CNN Model

HE Ruimin[1], ZHENG Kefeng[2], WEI Qinyang[1], ZHANG Xiaobin[2], ZHANG Jun[1], ZHU Yihang[2], ZHAO Yiying[2], GU Qing[2*]

(1. *Shengzhou Mosang High-tech Co. , Ltd.* , *Shaoxing* 312400, *China*;
2. *Institute of Digital Agriculture*, *Zhejiang Academy of Agricultural Sciences*, *Hangzhou* 310021, *China*)

Abstract: Factory-like rearing of silkworm (*Bombyx mori*) using artificial diet for all instars is a brand-new rearing mode of silkworm. Accurate feeding is one of the core technologies to save cost and increase efficiency in factory silkworm rearing. Automatic identification and counting of silkworm play a key role to realize accurate feeding. In this study, a machine vision system was used to obtain digital images of silkworms during main instars, and an improved Mask R-CNN model was proposed to detect the silkworms and residual artificial diet. The original Mask R-CNN was improved using the noise data of annotations by adding a pixel reweighting strategy and a bounding box fine-tuning strategy to the model frame. A more robust model was trained to improve the detection and segmentation abilities of silkworm and residual feed. Three different data augmentation methods were used to expand the training dataset. The influences of silkworm instars, data augmentation, and the overlap between silkworms on the model performance were evaluated. Then the improved Mask R-CNN was used to detect silkworms and residual feed. The *AP*50 (Average Precision at IoU=0.5) of the model for silkworm detection and segmentation were 0.790 and 0.795, respectively, and the detection accuracy was 96.83%. The detection and segmentation *AP*50 of residual feed were 0.641 and 0.653, respectively, and the detection accuracy was 87.71%. The model was deployed on the NVIDIA Jetson AGX Xavier development board with an average detection time of 1.32 s and a maximum detection time of 2.05 s for a image. The computational speed of the improved Mask R-CNN can meet the requirement of real-time detection of the moving unit of the silkworm box on the production line. The model trained by the fifth instar data showed a better performance on test data than the fourth instar model. The brightness enhancement method had the greatest contribution to the model performance as compared to the other data augmentation methods. The overlap between silkworms also negatively affected the performance of the model. This study can provide a core algorithm for the research and development of the accurate feeding information system and feeding device for factory silkworm rearing, which can improve the utilization rate of artificial diet and improve the production and management level of factory silkworm rearing.

Key words: silkworm; artificial diet; accurate feeding; machine vision; deep learning; mask R-CNN; noise data